ABSTRACTS

European Student Council Symposium 2016

3 Sept., The Hague, Netherlands

1. Subgrouping of pediatric medulloblastoma using an integrated analysis of MicroRNA-mRNA expression profile

Sivan Gershanov, Ariel University, Israel

Medulloblastoma (MB), the commonest malignant brain tumor of childhood, is divided into four tumor subgroups representing distinct clinical, biological and molecular entities. Subsequently, treatment should be designed according to the specific subgroup. MicroRNAs (miRNAs) are involved in carcinogenesis and tumor progression by regulating post-transcriptional gene expression. However, the miRNA-mRNA regulatory network in MB is far from being fully understood. The aim of the study is to identify novel miRNA subgroup biomarkers and their target mRNAs for rapid, specific and cost effective diagnosis by analyzing integrated mRNA-miRNA transcriptome sequencing from tumors. With this aim, integrated whole transcriptome mRNA and miRNA expression analysis was performed on primary tumor samples collected from 10 MB patients. 867 mature miRNAs were identified in at least a single MB sample, of them 462 were common to all 4 subgroups. 25 (2.5%) of all expressed miRNAs appeared to be significantly differentially expressed between the medulloblastoma subtypes (FDR<0.1). Namely, upregulation of hsa-miR-224-5p and hsa-miR-449c-5p was found exclusively among WNT, while downregulation of hsa-miR-135b-5p characterized SHH. Among groups 3 and 4, hsa-miR-20a-5p was upregulated or downregulated, respectively. RNA-seq from the same tumor samples identified 500 genes that vary between the four subtypes (q-value <0.05), among which 69 (13.8%) have anti-correlated miRNA-mRNA interactions with the 25 detected miRNA biomarkers. The predicted mRNAs targets of these miRNAs are associated with different signaling pathways, known to have a role in MB biology. Our study demonstrates that miRNAs are readily detectible and are highly specific to distinct MB subgroups. Understanding the involvement of miRNAs and their targets in MB related signaling pathways may improve diagnosis and advance the development of targeted treatment for MB.

2. From Predicting to Analyzing HIV-1 Resistance Towards Broadly Neutralizing Antibodies

Anna Feldmann, Max Planck Institute for Informatics, Germany

With around 37 million people living with HIV, an incidence rate of around 2 million per year and no cure in sight, clinicians have to face drug resistance emergences while the number of available drugs remains limited. Recently, combination therapy with broadly neutralizing antibodies (bNAbs) was introduced as a viable new option in antiretroviral treatment against HIV-1, that is capable of reducing viral load under detectable levels for up to 60 days in humanized mice and non-human primates. First clinical trials showed that already a single infusion of one bNAb, 3BNC117, is able to suppress successfully viremia in HIV-1 infected humans and even enhance the antibody responses of the individuals. However, the efficacy of this treatment is also affected by the emergence of resistant strains. Prior to the administration of an antiretroviral bNAb combination therapy to a patient, it has to be ensured that the patient's viral strains are susceptible to the particular bNAbs of the combination. So far, resistance to bNAbs can only be tested in expensive and time-consuming neutralization assays. For eleven different bNAbs, we propose a non-linear SVM-based model to predict the neutralization susceptibility of unseen viral strains to each of the bNAbs based on the viral envelope sequence (performance of up to 84% AUC). Because non-linear SVM classification results are often difficult to interpret, we offer different visualization techniques to improve the biological interpretability of the results using feature space visualization and motif logos. Learning the important binding sites of the bNAbs, the models are also biologically meaningful and useful for epitope recognition. Apart from the classifiers, we provide regression models that are more sensitive than the classifier approach by not using an empirical cutoff to distinguish susceptibility and resistance. Moreover, we confirmed a trend towards antibody resistance for the subtype B HIV-1 population and extended the analysis to the global HIV-1 population by predicting the neutralization sensitivity for around 36,000 HIV-1 sequences from the Los Alamos National Laboratory HIV Sequence Database.

3. TCGAbiolinks: An R/Bioconductor package for integrative analysis with TCGA data

Antonio Colaprico, Interuniversity Institute of Bioinformatics in Brussels (IB)², Belgium

The Cancer Genome Atlas (TCGA) research network has made public a large collection of clinical and molecular phenotypes of more than 10 000 tumor patients across 33 different tumor types. Using this cohort, TCGA has published over 20 marker papers detailing the genomic and epigenomic alterations associated with these tumor types. Although many important discoveries have been made by TCGA's research network, opportunities still exist to implement novel methods, thereby elucidating new biological pathways and diagnostic markers. However, mining the TCGA data presents several bioinformatics challenges, such as data retrieval and integration with clinical data and other molecular data types (e.g. RNA and DNA methylation). We developed an R/Bioconductor package called TCGAbiolinks to address these challenges and offer bioinformatics solutions by using a guided workflow to allow users to query, download and perform integrative analyses of TCGA data. We combined methods from computer science and statistics into the pipeline and incorporated methodologies developed in previous TCGA marker studies and in our own group. TCGAbiolinks downstream analysis can be divided into 1) supervised analysis, comprising differential expression analysis, enrichment analysis, and master regulator analysis or 2) unsupervised analysis,: comprising inference of gene regulatory network, cluster, classification, ROC, AUC, feature selection, and survival analysis. Using four different TCGA tumor types (Kidney, Brain, Breast and Colon) as examples, we provide case studies to illustrate examples of reproducibility, integrative analysis and utilization of different Bioconductor packages to advance and accelerate novel discoveries. [Availability] http://doi.org/10.1093/nar/gkv1507 Our package is freely available within the Bioconductor project at http://bioconductor.org/packages/TCGAbiolinks/. For detailed results see NAR's online paper about four case studies and for reproducible R codes in supplementary informations and related vignette. http://nar.oxfordjournals.org/content/suppl/2015/12/23/gkv1507.DC1/nar-03136-met-n-2015-File009.pdf TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages [http://f1000research.com/articles/5-1542/v1] [TCGAbiolinks's Applications] TCGAbiolinks it was recently used for section (4) mRNA Expression and (5) DNA methylation profiling in last TCGA's marker paper published in early 2016. See citation in supplementary informations. Ceccarelli et al, Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma, Cell 164 Issue 3: p550-563, 2016 https://tcga-data.nci.nih.gov/docs/publications/ http://www.sciencedirect.com/science/article/pii/S009286741501692X http://www.sciencedirect.com/science/MiamiMultiMediaURL/1-s2.0-S009286741501692X/1-s2.0-S009286741501692X-mmc1.pdf/272196/html/S009286741501692X/d3371d49bef3810100e52cf6ec732f74/mmc1.pdf [Bioconductor's links] https://www.bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html https://www.bioconductor.org/packages/release/bioc/vignettes/TCGAbiolinks/inst/doc/tcgaBiolinks.html http://bioconductor.org/packages/stats/bioc/TCGAbiolinks.html

4. Whole genome sequencing of a dried blood spot reveals an atypical T. gambiense

Bart Cuypers, Institute of Tropical Medicine, Belgium

African trypanosomes are lysed by the human serum protein apolipoprotein L1 (ApoL1), except Trypanosoma brucei gambiense and T. b. rhodesiense, which cause human African trypanosomiasis. These two subspecies can resist human ApoL1 because they express the serum resistance proteins TgsGP and SRA, respectively. Here we report a complex interplay between trypanosomes and an ApoL1 variant, revealing important insights into innate human immunity against these parasites. Using whole genome sequencing, we characterized an atypical T. b. gambiense infection in a patient in Ghana. Total DNA was extracted from a dried blood spot of the patient and sequenced on an Illumina MiSeq. We performed a genome-wide SNP analysis of the trypanosome and Western Africa T. b. brucei and T. b. gambiense reference genomes. After stringent quality filtering, a maximum likelihood tree was created with RAxML. Phylogenetic analysis showed that the infecting trypanosome has diverged from the classical T. b. gambiense strains. Additionally, the genomic analysis revealed that the trypanosome lacks the T. b. gambiense specific TgsGP defense mechanism against human serum, and should thus not be infective to man. By sequencing the ApoL1 gene of the patient and subsequent in vitro mutagenesis experiments, we demonstrated that a homozygous missense substitution (N264K) in the membrane-addressing domain of the patient's ApoL1 knocks down the trypanolytic activity, allowing the TgsGP-lacking trypanosome to avoid ApoL1-mediated immunity. Our data suggest that populations with high frequencies of the homozygous N264K ApoL1 variant may be at increased risk of contracting human African trypanosomiasis.

5. A quick and flexible transcriptomic feature quantification framework on the cloud

Andrian Yang, Victor Chang Cardiac Research Institute, Australia

Major advancement in single-cell capture technology has resulted in the increasing interest in single-cell level studies, particularly in the field of transcriptomics. Current tools designed for transcriptomic analysis are unable to efficiently handle this increasingly large volume of sequencing data generated. To tackle this problem, we have implemented a cloud-based framework for the simultaneous processing of large-scale transcriptomic data. The pipeline utilises state-of-the-art Big Data technology of Apache Hadoop, a MapReduce framework, and Apache Spark, a general purpose data analytics engine, to perform massively parallel alignment and feature quantification analysis of transcriptomic data on a cloud-computing environment which can be scaled to meet user requirements. The default pipeline makes use of STAR for sequence alignment and featureCount for feature quantification. Nonetheless, the pipeline is customisable in terms of choice of parameter and tools for alignment and feature quantification. Our framework also performs RNA-seq data quality control using Picard. We evaluated the performance of the pipeline using a public single-cell mouse RNAseq dataset (869 samples, 1.28T bases) on a 10 node Amazon Elastic MapReduce cluster (320 cores, 2.21TB RAM). The analysis was completed in 6.75 hours, which is 4.3x faster compared to performing the same analysis on an equivalent single computing resource. The pipeline offers the use of low-cost spot instances, providing a saving of 3.32x (US$65.10 spot vs US$216.30 on-demand) for the analysis performed.

6. Data fusion in drug target interaction prediction and drug repositioning

Daniele Parisi, KULeuven, Belgium

Although the repositioning of existing drugs for new indications can potentially avoid expensive costs associated with early-stage testing of the hit compounds, however no perfect computational software for prediction is available yet. In fact, each approach considers just specific typology of data, disregarding other important information. With my project I try to improve the prediction power in drug repositioning and drug design, by fusing different approaches (network-based, ligand-based, and structure-based) by a multi-scale data-integration tool, MACAU. So far I have computed a large-scale drug-target interaction analysis, comparing 60.000 interaction complex profiles from PDB crystals, by using a pharmacophore fingerprint based approach (KRIPO), resulting in 19 millions of Tanimoto coefficient similarity scores. Currently, I am integrating my data with other affinity information from a large scale computational experiment of interaction (91 millions of docking scores), and experiementally obtained affinity data from CHEMBL (IC50 and Ki of drug-target couples). Furthermore, I'm using those structure-based data for helping the designing of a new class of immunosuppressant drugs targeting B-cells by fragment-based drug design and drug repositioning.

7. XGSA: A statistical method for cross-species gene set analysis

Djordje Djordjevic, Victor Chang Cardiac Research Institute, Australia

Gene set analysis (GSA) is a powerful tool for determining whether a set of genes is enriched for genes in known pathways or gene ontology terms. Current GSA methods do not facilitate comparing gene sets across different organisms as they do not explicitly deal with homology mapping between species. There lacks a systematic investigation about the effect of complex gene homology on cross-species GSA. In this work, we show that not accounting for complex homology when performing cross-species GSA leads to false positive biases. We propose XGSA, that explicitly takes homology mapping into consideration when doing cross-species GSA. Simulation experiments confirm that XGSA can avoid false positive discoveries, while maintaining good statistical power compared to other ad-hoc approaches for cross-species gene set analysis. We demonstrate the effectiveness of XGSA with case studies that aim to discover conserved or species-specific molecular pathways involved in social challenge and vertebrate appendage regeneration.

8. Discovering molecular signatures of extreme physiology using African mole-rats

Ole Eigenbrod, Max Delbrueck Center for Molecular Medicine in the Helmholtz Association, Germany

The African mole-rats (Bathyergidae) are a family of subterranean rodents with very unusual physiological traits for mammals. The most famous member of African mole-rats is the naked mole-rat (Heterocephalus glaber), which shows several extraordinary phenotypes like poikilothermy, extreme longevity, cancer resistance and extreme adaptation to low oxygen environments. Additionally, the naked mole-rat and some other Bathyergidae species are insensitive to several noxious substances or algogens (e.g. acid, capsaicin, or mustard oil) [Park et al., PLOS Biology 2008]. This study focuses on understanding the sensory phenotypes of at least 8 African mole-rat species, as these closely related species show different patterns of insensitivity to noxious substances. Recently, a sequence motif in the NaV1.7 ion channel of the naked mole-rat was found to be directly connected to its acid insensitivity [Smith et al., Science 2011]. We sequenced poly-A selected mRNA from multiple tissues of 8 African mole-rat species. As there are no annotated genomes available for most of the species, we performed de-novo transcriptome assembly to obtain the protein-coding sequences. We developed a bioinformatic workflow to annotate putatively coding transcripts and exclude contaminating or falsely assembled sequences and chimeras. Using this approach, we were able to identify more than 10,000 unique protein-coding transcripts per species. We also directly compared the protein-coding sequences and transcript levels across species boundaries. This approach allows a multivariate analysis of the relationship between gene expression level, sequence variation and extreme phenotypes across this rodent family.

9. An integrated transcriptomics and proteomics study of embryonic stem cell differentiation

Patrick van den Berg, Leiden University, The Netherlands

Embryonic stem cells (ESCs) can be differentiated into all cell types of the adult body. In vitro differentiation of ESCs has therefore been used extensively as a model for embryonic development and it is critical for applications of ESCs in regenerative medicine and disease modeling. To differentiate ESCs into well-defined cell types, precise manipulation of gene expression is necessary. The majority of existing work has focused on transcriptional regulation of expression. Here, we study gene regulation at the level of protein turnover (translation and degradation) to discover novel ways to control ESC differentiation. In particular, we extracted mRNA and protein during retinoic acid induced differentiation of mouse ESCs. mRNA and protein abundance were then quantified by RNA sequencing and mass spectrometry, respectively. The measurement of 10 samples during a 96 h differentiation time course allowed us to follow the expression dynamics with unprecedented temporal resolution. We have developed a statistical model that identifies genes that are differentially regulated at the mRNA and protein level. After validation of the identified candidate genes we will unravel the general mechanisms that underlie their regulation.

10. Analysis of the structure, function and evolution of caleosins: a family of multifunctional eukaryotic proteins

Farzana Rahman, University of South Wales, United Kingdom

The multifunctional calcium-binding proteins termed caleosins occur almost ubiquitously in two distinct eukaryotic clades, namely Viridiplantae and Fungi. The evolutionary pattern of caleosin gene occurrence is not consistent their descent from a common ancestor because the Fungi, along with animals and many protists, are members of the Opisthokonta, while the Viridiplantae are derived from unrelated green algal predecessors. This suggests that the caleosin genes may have originated in one of the current clades via by horizontal gene transfer from the other. We have studied the variation in caleosin gene and protein sequences across a comprehensive range of plant and fungal species utilising computational methods and pipelines to understand the structure and function of these proteins in detail. Protein structure predictions suggest that the calcium-binding and EF hand domains are widely conserved across species, while there is considerable variation in the predicted loop region of the structure. While the biological functions of studied proteins have yet to be determined in detail it is clear that these proteins have several subcellular locations and participate in a range of physiological processes in both plants and fungi, including acting as peroxygenases. One of the most important of these roles appears to be in responses to a range biotic and abiotic stresses, including plant-fungal interactions. In this research we describe additional studies that have been carried out to shed light on the origin and functions of this intriguing group of proteins.

11. Three-dimensional chromatin looping predicted by CTCF motifs and ChIP-seq signals

Jonas Ibn-Salem, Johannes Gutenberg University Mainz, Germany

In eukaryotes, the transcription of gene is regulated by transcription factors (TF) which bind to gene promoters or distal regulatory elements such as enhancers that can be over 1Mb apart from the regulated gene. The regulation over such distances is facilitated by chromatin looping that bring the enhancer region into close physical proximity with the promoter. While TF binding sites can be detected genome-wide by ChIP-seq experiments it is difficult to associated distal binding sites to regulated genes without information of chromatin looping. Recent experimental techniques such as Hi-C or ChIA-PET detect chromatin lopping events genome-wide but are experimentally more elaborate and have either limited resolution or can be applied only for specific conditions. However, the formaldehyde treatment in the ChIP-seq protocol result in cross-linking of proteins with DNA. Therefore also regions that are not directly bound by the targeted TF but in close physical proximity through chromatin looping are purified in the immunoprecipitation step. If the TF binding site is in proximity with other gnomic regions such as promoters, we expect ChIP-seq signals at both loop anchor regions. It was shown that chromatin loops are mediated by CTCF proteins which recognize sequence motif in convergent orientation at loop anchors. We hypothesized that we can use the position and shape of ChIP-seq signals at convergent CTCF motif sites to predict whether they perform looping interactions or not. We used measured loops from high resolution Hi-C in human GM12878 cells and ChIP-seq data sets of four TFs. We aligned loop anchors on convergent CTCF motif and compared the similarity of ChIP-seq profiles at anchor pairs to randomly permuted anchor pairs by calculating the Euclidean distance of ChIP-seq coverage vectors. All TFs profiles show a significantly higher similarity at actual loop pairs than the permuted control anchor pairs. These preliminary results suggest that chromatin looping events can be predicted by ChIP-seq and sequence motif information alone at base-pair resolution. The predicted loops can be used to associate TF binding sites to regulated genes. Furthermore, application to hundreds of published ChIP-seq profiles will give insights which factors are functionally involved in chromatin looping.

12. Functional impact of genomic rearrangements on chromatin organization and transcriptional regulation

Sascha Meiers, European Molecular Biology Laboratory, Germany

With chromatin conformation capture-based techniques such as Hi-C it has become possible to study the interaction between cis regulatory elements in the genome (enhancers, promoters, etc.) at a genome-wide scale. Yet our understanding of how these interactions form and under which circumstances they regulate gene expression is only rudimentary. Recent studies investigated somatic chromosomal aberrations or used CRSIPR/Cas9 to edit key regions such as boundaries of topologically associated domains to understand the functional consequences of rearrangements. However, those results remain limited to few exemplary cases. In this ongoing work we used highly rearranged balancer chromosomes in Drosophila melanogaster as a genome-wide model for large-scale genomic rearrangements to investigate how the linear structure of the genome influences chromatin organization and gene expression. We analyzed expression in adult flies as well as embryos and compared the rearranged chromosomes to their normal state in a heterozygous cross, which intrinsically normalizes for trans regulatory effects. Surprisingly, initial results suggest little drastic effects, with some changes in the larger chromatin interaction landscape and hundreds of genes showing moderate differential expression between the two haplotypes. While analysis is still ongoing, we expect this model to yield unique insights into the interplay between chromatin topology and transcriptional regulation in cis.

13. Utilizing the Benford law for unravelling tissue specificity

Deepak Karthik, Ariel, Israel

The reduction in sequencing costs has led to an unprecedented trove of gene expression data from diverse biological systems. Subsequently, principles from other disciplines such as the Benford law, which can be properly judged only in data-rich systems, can now be examined on this high-throughput transcriptomic information. The Benford law states that in numerical data, the proportion of numbers beginning in any given digit is not uniform but rather skewed, with 1 being the most common digit and 9 the rarest. Here we demonstrate that digital gene expression data has a Benford-like distribution when observing an entire gene set. This phenomenon was conserved in a wide range of biological tissues and developmental conditions. However, when obedience to the Benford law is calculated for individual expressed genes across thousands of cells, genes that best and least adhere to the law are enriched with tissue specific or cell maintenance descriptors, respectively. Surprisingly, a positive correlation was found between the obedience a gene exhibits to the Benford law and its expression level, despite the former being calculated solely according to first digit frequency while totally ignoring the expression value itself. These results demonstrate the applicability and potential predictability of the Benford law for gleaning biological insight from simple count data.

14. A study of normal CNV variations in Israeli population

Pola Smirin-Yosef, Ariel University, Israel

The Israeli population is composed of a collection of diverse ethnic groups. Each group shares specific genetic variations that passed from its common ancestors throughout the generations. Together with pathogenic events, non-pathogenic polymorphism happen to occur in ancestors, subsequently spread into the restricted genomic pool of its descendants. Providing a comprehensive data resource of non-pathogenic CNVs in the Israeli population pregnancies in order to characterize ethnic-specific polymorphism may greatly contribute to the routine genetic counseling done by the geneticists on a daily basis. Chromosomal Microarray Array (CMA) has had a high impact in clinical diagnostics, leading to the discovery of new genomic disorders, and has become an indispensable tool for routine molecular and cytogenetic testing. CMA is a first line diagnostic test for individuals with developmental disabilities, dysmorphic features and congenital malformations as well as fetuses with congenital malformations and abnormal growth. Here we apply a data mining approach on the results of CMA testing performed at the Raphael Recanati Genetic Institute, contains around 9000 tests from individuals, fetuses with clinical abnormalities, and in fetuses from low-risk pregnancies. The use of an extracted ethnicity-based genetic information, in order to detect ethnic-specific CNV polymorphism in the Israeli population will allow geneticists to distinguish between relevant pathogenic genomic aberrations from benign ethnicity-related variations.

15. AmyloGram: a novel predictor of amyloidogenicity

Michał Burdukiewicz, University of Wrocław, Poland

Background: Amyloids are proteins associated with the number of clinical disorders (e.g., Alzheimer’s, Creutzfeldt-Jakob’s and Huntington’s diseases). Despite their diversity, all amyloid proteins can undergo aggregation initiated by 6- to 15-residue segments called hot spots. Henceforth, amyloids form unique, zipper-like β-structures, which are often harmful. To find the patterns defining the hot spots, we developed our novel predictor of amyloidogenicity AmyloGram, based on random forests. We trained it using short motifs (n-grams) extracted from amyloid and non-amyloid peptides collected in the AmyLoad database. Description: Peptide data were represented by various amino acid physicochemical properties. We tested 524 284 random forest predictors, each employing reduced amino acid alphabet based on a different combination of the physicochemical properties of residues. As a result, we identified the reduced alphabet providing the best discrimination between amyloids and non-amyloids, which was based on the hydrophobicity index, polarizability parameter, β-sheet propensity and average flexibility. Three first features are well-known factors in amyloidogenicity, but the role of the last one in this process was previously unknown. Most of the predictors based on reduced amino acid alphabet outperformed a random forest trained on the full amino acid alphabet confirming our assumption on the role of more general amino acid properties. During analysis we also found 65 n-grams that are most relevant to the discrimination of amyloid and non-amyloid sequences, 15 motifs were independently confirmed experimentally elsewhere. The best-performing predictor, AmyloGram, was benchmarked against the most popular tools for amyloid peptides detection using an external dataset. Our software obtained the highest values of performance measures (Area Under the Curve: 0.90, Matthews correlation coefficient: 0.63). Conclusions: The n-gram analysis not only confirmed that amyloidogenicity depends on the general physicochemical properties of proteins, but also revealed which features are the most relevant to the initiation of amyloid aggregation. In addition, our framework identified amyloidogenicity-related amino acid motifs, which were partially confirmed experimentally. Aside from creation of the interpretative model of amyloidogenicity, we also established the accurate predictor of amyloids, AmyloGram, which is available as a web-server: www.smorfland.uni.wroc.pl/amylogram/.

16. Testing for association between RNA-Seq and high-dimensional data

Armin Rauschenberger, VU University Medical Center, The Netherlands

Background It is often of interest to know how RNA-Seq gene expression data is associated with other complete molecular profiles, such as single nucleotide polymorphisms (SNPs) and DNA methylation. However, no parametric statistical test could so far do this, due to the overdispersion characteristic of RNA-Seq data. Description We will develop just such a test. It is based on a regression model, where the expression of one gene is the response variable, and multiple probes from one molecular profile are the covariates. The response is discrete and typically shows high variability, and the number of covariates may exceed the sample size. Therefore we assume that the response follows a negative binomial distribution and that the effects of the covariates on the response are realisations of a random variable with zero mean and an unknown variance. Under the null hypothesis no covariate influences the response, whilst under the alternative hypothesis at least one of them does. We use the score of the variance parameter as a test statistic, and obtain p-values via permutation. We show in a simulation study that the proposed test successfully deals with overdispersion of the response and with high dimensional settings. It is powerful at finding effects and maintains the false positive rate. Furthermore, the test statistic can be decomposed into contributions of individual samples or covariates to the test statistic. Finally, the test can be extended to test association between RNA-Seq and multiple molecular profiles (copy number variation, loss of heterozygosity, microRNA expression) simultaneously. In two applications we detect genetic and epigenetic alterations that may affect gene expression. Conclusions We propose a parametric test for associations between a count variable (such as RNA-Seq) and large sets of quantitative or binary variables. The test is implemented in the R package globalSeq from Bioconductor. Rauschenberger A, Jonker MA, van de Wiel MA and Menezes RX (2016). “Testing for association between RNA-Seq and high-dimensional data.” BMC Bioinformatics, 17:118.

17. Klinefelter syndrome comorbidities induced by increased X gene dosage and altered protein interactome activity

Francesco Russo, National Research Council of Italy and Novo Nordisk Foundation Center for Protein Research (University of Copenhagen), Denmark

Klinefelter syndrome (KS) (47,XXY) is the most common male sex chromosome aneuploidy. Diagnosis and clinical supervision remain a challenge due to varying presentation and insufficient characterization of the syndrome. Here we present a study combining health data-driven epidemiology and molecular level systems biology to improve the understanding of KS and the molecular interplay influencing its comorbidities. Using health registry data from the entire Danish population covering 6.8 million patients a total of 78 overrepresented comorbidities were identified from Danish hospital patient records. The extracted comorbidities included both clinically well-known (e.g. infertility and osteoporosis) and still less established KS comorbidities (e.g. pituitary gland hypofunction and dental caries). Three approaches were applied to identify key underlying molecular players in the KS comorbidities: (A) Differential expression analysis and identification of coexpressed modules using data generated on peripheral blood, (B) Identification of central hubs in a KS comorbidity network based on known disease proteins and their protein-protein interactions, and (C) Identification of dosage perturbed protein complexes in the KS comorbidity network. Together these approaches pointed to novel aspects of X-chromosome related mechanisms, including perturbed Cytokine-cytokine interaction and Jak-Stat pathways, immune system alterations and disturbed functionality of leptin and erythropoietin signalling in KS. This work presents an extended epidemiological study that links KS comorbidities to the molecular level and identify potential causal players in the disease biology underlying the identified comorbidities.

18. Unraveling the impact of the human cytomegalovirus-encoded chemokine receptor US28 on Wnt/β-catenin signaling

Anne-Merel van der Drift, Centre for Integrative Bioinformatics (IBIVU), VU University Amsterdam, Amsterdam, The Netherlands

The human cytomegalovirus (HCMV), encoding the constitutively active chemokine receptor US28, is suggested to act as an oncomodulator. US28 activates many different signaling pathways involved in proliferation including Wnt/β-catenin signaling. US28 stimulation leads to β-catenin accumulation by signaling via the RhoA/ROCK pathway. Further, ROCK is predicted to activate HGF/MET signaling, leading to the release of β-catenin from the adhesion complex, hence β-catenin accumulation independent from Wnt/β-catenin signaling. However, the exact regulatory role of ROCK is still unknown. Here, we hypothesize that ROCK activates multiple proteins to regulate this process: i) direct stimulation via AKT, ii) negative feedback via PTEN and iii) feedforward stimulation via the adhesion complex. We built a Petri net model of US28/β-catenin signaling based on the known mechanisms together with our hypothesized mechanisms. Simulations of our model recapitulated experimental observations from literature: i) dose-dependent US28/β-catenin signaling, ii) higher β-catenin accumulation of US28 stimulation compared to Wnt stimulation and iii) independent β-catenin signaling via Wnt and US28. Further, simulations of our validated model with PTEN inhibition, i.e. inhibition of the negative feedback, showed maximal β-catenin accumulation for all levels of US28 stimulation. Our validated model proposes a hypothetical mechanism for US28/β-catenin signaling, where ROCK is an important regulator. Further, US28 signaling leads to the release of β-catenin from the adhesion complex resulting in independent signaling from Wnt. The release of β-catenin bound to the adhesion complex dissociates the cell-cell adhesion which might be beneficial for migration of cancerous cells. Moreover, PTEN is important for the negative feedback and insures a dose-dependent response. The highly oncogenic PTEN could be involved in the oncomodulatory effect of the human cytomegalovirus.

19. Robust In-Silico identification of sequenced Cancer Cell Lines

Raik Otto, Humboldt-Universität zu Berlin, Germany

Cancer cell lines are a pivotal tool for cancer researchers. However, cancer cell lines are prone to critical errors such as misidentification and cross-contamination which have reportedly caused severe setbacks. Established cancer cell line identification methods compare genotype characteristics obtained during specific experiments (e.g. SNP arrays); characteristic genotype properties of the to-be-identified sample (the query) are matched against the same characteristics properties of the known samples (the references). If a match shows a significant similarity to a reference sample, the query is identified as the reference sample. Such characteristic genotype information can also be derived from NGS data. A query can be identified when the characteristic genotype properties were obtained from Next-generation sequencing of the query and a subsequent comparison to a NGS reference. However, results from different NGS technologies, algorithms and sequencing-approaches, e.g. whole-exome or panel-sequencing, are inherently challenging to compare. SNP-zygosity matching and tandem repeat-counting on such data is in general unreliable due to non-covered loci, SNP-filtering, and zygosity-call divergence caused by differing algorithmic ploidy-settings. Here, we present the Uniquorn method that reliably identifies cancer cell line samples based on NGS genotyping data across different technologies, algorithms, filter-settings and covered loci. Uniquorn compares the query to all references and computes a p-value for the likelihood that an overlap in observed genomic variants is due to chance. Uniquorn was benchmark by cross-identifying 1989 cancer cell line sequencing samples: sensitivity amounted to 96% and specificity to 99%. The R-BioConductor package Uniquorn and the benchmark setup are freely available.

20. Open Source Development Success through collaboration; SmartR in tranSMART

Jochem Bijlard, The Hyve, The Netherlands

Summary: Through collaboration on open source software top pharmaceutical companies, academic researchers and professional software development companies have proven success in enhancing the tranSMART platform with an interactive and commercial quality visualization platform. tranSMART is an open source translational research platform used by academic researchers and pharmaceutical companies around the world. The tranSMART Foundation, supported by many of these users, guards the quality of the platform by setting code standards and encouraging collaboration. The Innovative Medicines Initiative (IMI) project eTRIKS is the result of a collaboration between 17 different academic and industrial partners. Each combining their strengths in the development of a platform and services for data staging, exploration and use in translational research. Within eTRIKS one of the academic partners, University of Luxembourg, developed a new visualisation platform for within tranSMART, called SmartR. SmartR is aimed to provide a highly dynamic and interactive way of visualizing and analyzing data within tranSMART. Using recent web technologies it generates interactive analytics within the web browser rather than making use of static images generated by R. Academic and industrial environments put different constraints and requirements on software development. Where academic developments are focussed on proving the validity of a novel innovation, software for industrial research needs to be scalable and reliable. Within separate development projects and hackathons the pharmaceutical companies Pfizer and Sanofi have sponsored the open source bioinformatics software company The Hyve to work with the original developer to upgrade the SmartR visualization platform to be of commercial quality in code and analysis algorithms and allow for easy extension with more workflows. By leveraging the trust built in the open source community these competing companies have involved each other in their projects building towards a common goal. Active collaboration is still underway to release the enhanced SmartR as a default plugin with the 16.2 version of tranSMART, which will be released in the second half of 2016.

21. Open Source Development Success through collaboration: Contributions to cBioPortal

Jochem Bijlard, The Hyve, The Netherlands

Summary: Through collaboration on open source software, pharma, commercial software development and cancer research institutions have proven successful in enhancing the cBioPortal platform by optimizing and extending it with new features. Approximately one year ago the popular cBioPortal for Cancer Genomics was made open source. In this last year its development community has grown and the platform has been extended with many new features. Here we detail some of the contributions The Hyve (Utrecht) has made to the platform, in collaboration with Dana Farber Cancer Institute (Boston), Memorial Sloan Kettering Cancer Center (New York) and Boehringer Ingelheim (BI RCV). The contributions can roughly be divided into three categories: (1) improvement of the data loading pipeline, (2) new data analysis features, and (3) optimizations of the front end. In the data loading pipeline we have introduced a strict separation between the validation step and the loading step. This “separation of concerns” design principle makes the code easier to understand and maintain and simplifies the process of adding new datasets to a local cBioPortal installation. Special effort was spent on making the validator easy to use, which is exemplified by clearer error messages and the generation of a HTML validation report. In the front end we added a whole new pancancer view for studies comprising multiple cancer types, added new query options in the Study overview page and added new visualizations to the query results

page to support better enrichment analysis of expression (mRNA, Proteins) and cooccurrence (copy number, mutations). We have also implemented integration documentation from the Wiki or Git, and made the portal more customizable (logo, headers, news and FAQ), which is very important for open source software. Last but not least, we have optimized the loading times of the portal to be able to host larger studies, focusing on the most used pages in the application. In the query results page we have successfully shortened the loading times of various analyses.

22. Improving potato breeding with computational and functional genomics

Konrad Zych, Groningen Bioinformatics Centre, University of Groningen, The Netherlands

Potato is one of the most important food crops. Potato is an outbred tetraploid plant making its breeding time-consuming and cumbersome. Including genetic markers in the selection process could greatly improve potato breeding. This approach was successfully used in selection for few monogenic traits (e.g. resistance to Phytopthora infestans). In our study we developed markers for reliable screening for multigenic quality traits like color after frying. We created a large potato population, consisting of two experimental crosses and a panel of cultivars and breeding clones. We performed RNA-Seq on the parents of the crosses in order to extract SNPs, from which we created a 60,000 SNP array. We used this array to genotype our population. We extended the mixture models based genotype calling of fitTetra (Voorrips et al, 2011). We used RNA-Seq data to obtain starting values for the algorithm increasing accuracy of the calling. The resulting genotypes were used together with multi-year high quality phenotypes in association studies. Using multiple levels of correction for population structure and environmental variance and multiple-marker association analysis we elucidated new markers for complex potato quality phenotypes. With our improved algorithm we were able to salvage 20% more high quality SNPs and filter out the lower quality SNPs. As a result, we created one of the most comprehensive genomic resources for potato with more than 30,000 SNPs measured in more than 1,500 samples. Association analysis resulted in a set of markers that could be used by the companies to extend their breeding scheme.

23. Investigating allosteric coupling using fuzzy ligands

Susanne Hermans, Heinrich-Heine University, Germany

The identification of novel allosteric pockets suitable as drug targets is complicated by the variety of allosteric mechanisms. In the context of dynamically dominated allostery, i.e. in the absence of conformational change, [1] we present a new approach to probe allosteric signaling through proteins by generating fuzzy ligands as surrogates for “true” molecular ligands. The approach was applied to ensembles generated by MD simulations of the crystal structures of LFA-1, CAP and TRAP proteins. PocketAnalyzerPCA [2] was used to detect putative allosteric sites, and fuzzy ligands were generated for each pocket along the ensemble. Finally, the Constraint Network Analysis (CNA) software was applied, [3] which performs a global and local stability analyses. The altered stability characteristics upon binding of the fuzzy ligand and the co-crystallized ligand are compared to validate the approach. Remarkably, fuzzy ligands almost perfectly reproduce the perturbation observed for the co-crystallized ligand. Our fuzzy ligand approach can thus be used to probe allosteric effects without any information about a natural ligand. Analyzing unexplored pockets with fuzzy ligands can be a promising step towards identifying novel drug targets. [1] A. Cooper, D.T. Dryden, Eur. Biophys. J., 1984, 11, 103-109 [2] I.R. Craig, C. Pfleger, H. Gohlke, J.W. Essex, K. Spiegel, J. Chem. Inf. Model., 2011, 51, 2666-2679 [3] C. Pfleger, P.C. Rathi, D.L. Klein, S. Radestock, H. Gohlke, J. Chem. Inf. Model., 2013, 53, 1007-1015

24. Improving the prediction of processed pseudogenes in human

Sweta Talyan, Johannes Gutenberg University, Germany

Pseudogenes are extant genomic loci that are quite similar to their parental, functional genes, but cannot be translated into functional proteins because of deleterious mutations like frameshift disruptions or premature stop codons. Pseudogenes are classified depending on their biogenesis mechanism: retrotransposition, DNA duplication and gene decay respectively as processed, duplicated and unitary. Duplicated pseudogenes can maintain the parental gene structure and all regulatory regions, processed pseudogenes are not able to retain neither the 5’ upstream regulatory regions nor the introns. Recent studies confirm the tissue specific, transcripional activity of more than 13% of all human pseudogenes. For some of those, functional regulatory roles have been found, including being causative of diseases. Currently, psiDr/GENCODE is the standard repository of pseudogene annotations. It is based on Pseudopipe and Retrofinder prediction methods followed by HAVANA manual curations. These methods of ab-initio pseudogene detection and classification were developed at an early stage of the human genome annotation, when little sequencing information from human and other organisms was available: Pseudopipe (Zhang et al. 2003, 2006, Zhang and 2004), Retrofinder (Baertsch et al. 2008) and the method from Torrents et al. (2003). These methods are still the norm and rely mostly on homology. They mainly differ in the parental gene query representative, the use of parameters and thresholds, and the incorporation of amino acid substitution replacement (Ka/Ks) measurements. In the wake of data availability, better pseudogenes annotation is nowadays necessary for human and some other species. Towards this aim, we aim to develop a novel method for pseudogene genome-wide prediction specially processed pseudogenes that takes advantage on information provided by the annotation on all the genomes sequenced till now, which will improve the current pseudogene annotation and classification. We aim to set up a new standard on pseudogene annotations.

25. Nucleus specific expression in the multi-nuclear mushroom Agaricus bisporus

Thies Gehrmann, Delft University of Technology, The Netherlands

Background: Many fungi, whose importance in our environment and industries can hardly be understated, have more than one nucleus per cell. The average cell in the cultivated white button mushroom, Agaricus bisporus, contains six nuclei, each originating from one of the two parental nuclei, referred to as the homokaryons of A. bisporus. Genes therefore exist in two different forms, called karyolleles, once in each homokaryon. The two homokaryons of A. bisporus A15 are called P1 and P2. We examine for the first time, the spatiotemporal karyollele specific expression of genes in a fungus. Methods: Using gene predictions for the genome sequences of both the P1 and P2 homokaryons, we identify karyollele pairs. Unique markers that distinguish them are discovered and quantified in RNA-seq data from different tissues throughout the development of the mushroom. Together with functional predictions, we examine the role of differentially expressed karyolleles. Results: We find that the P1 and P2 nuclei are differentially active in different mushroom tissues throughout development. In fact, P1 nuclei of different samples are transcriptomically more similar to each other than to P2 nuclei in the same sample. Furthermore, we find that chromosomes in the different nuclei are also differentially active. The regulation occurs at the gene level, rather than at a chromosome or nuclear level. This is indicated by neighbouring karyolleles on the same chromosome which are upregulated in different nuclei. We find 412 differentially expressed genes throughout development. These genes represent a large variety of functionality, including metabolism and secreted proteins. Differential karyollele expression in mushroom tissues, where P1 and P2 upregulate about the same number of genes, changes in compost where P2 upregulates more genes than P1. Conclusion: That near-identical genes in the different homokaryons are differentially active in response to the same external conditions, reveals a complex regulation between nuclei, and highlights their importance as individual protein products. An improved understanding of this behaviour is necessary to understand the regulatory mechanisms that control the differential regulation, and to improve the machinery for industrial exploitation.

26. RepeatsDB 2.0: improved annotation, classification and visualization of repeat proteins

Lisanna Paladin, Università degli Studi di Padova, Italy

Repeat proteins are a widespread class of non-globular proteins carrying fundamental functions and largely involved in several diseases. Their abundance in eukaryotes (one in three proteins in mammals according to recent estimates) suggests an important role in the evolution of complex organisms. RepeatsDB database was developed in 2013 to provide a resource for tandem repeat proteins. It contains more than 10,000 structures collected from the Protein Data Bank (PDB) and predicted to be repeated. It features also a detailed and manually curated structural characterization covering the 3% of the entries, annotated with the exact position of the repetitive elements (units). Taking advantage of the Structure Repeat Unit Library (SRUL) obtained through this manual curation, we developed and published a method, Repeat Protein Unit Predictor (ReUPred), for the fast automatic prediction of repeat units and repeat classification. Here we present a new release of the database, RepeatsDB 2.0, based on ReUPred scanning of the whole PDB and featuring extensive manual validation of the entries. It includes more annotations, an improved classification, a new search engine and a new web interface. The new search engine allows complex queries across Uniprot, PDB and repeat region features and including logical operators. In addition, it is now possible to retrieve entries by functional features (GO terms), domains (Pfam) and disease annotations (OMIM). The new web interface allows the view of repeat units mapped on PDB sequence and structure, together with secondary structure and fold information, Pfam domain assignment and reported SNPs and modifications. Moreover, a new classification level has been added to include functional relationship among different repeat units based on Pfam domain assignments. The new classification has been implemented as an independent layer on top of the existing structural features. Finally, a feedback service has been included for each entry with the aim of collecting user provided annotation, thereby improving RepeatsDB annotation.

27. Quantitative Protein Expression, In-silico

Anton Semenchenko, CEFETMG, Brazil

Background The development of the hybrid model of mRNA translation is be presented. The computational system for quantitative protein synthesis studies is constructed by the combination of the 3D cellular automata and agent-based simulation of the amino acid elongation. Markov chain representation of the biochemical processes orchestrated by the ribosome is the fundamental element of the model. Description The model of the mRNA translation is validated by the experimental data from Luciferase production in the cell-free system. The ease of use, flexibility and rigor of the agent-based technique is demonstrated by the evaluation of the two hypotheses regarding the inhibition mechanisms of the edeine antibiotic. The analysis and visualization of the polysome "nursery", ribosome traffic jams in 3D environment are presented. The long range interactions and other stochastic properties of the mRNA-ribosome systems are investigated using the Hurst exponent. Also, the influence of the EF-4 (LepA) is quantitatively described using computer simulation validated by the experimental data. Conclusions The presented computational system is capable to carry-out the investigation of the non-stationary regimes and spacial properties of the polysomes. The application of this model is illustrated by the virtualization of the cell-free protein expression kits. Moreover, the opportunities of the future research of the mRNA translation and protein formation are discussed in detail.

28. Genome-Wide Characterization of Folate Transporter Proteins for Eukaryotic Pathogens; Searching for the Next-Generation Antifolate Chemotherapy

Benson Otarigho, Department of Biological Science, Edo University Iyamho, Nigeria

Protozoan parasites, which cause infectious diseases, such as malaria, sleeping sickness, Chagas' disease and leishmaniasis are a global threat. Increased morbidity and mortality, lack of vaccines and the rapid spread of drug-resistant strains, call for specific novel strategies to combat these parasites. Parasite genome sequencing projects have assisted in identifying tractable protein drug targets such as transporters for folate, which might represent attractive therapeutic targets. Therefore, this study investigated the whole eukaryotic pathogen genomes for genes encoding homologues of proteins that can mediate the transportation of folate. We analysed about 200 genomes of pathogenic protozoan parasites in EupathDB for genes encoding homologues proteins that mediate folate transportation: folate-binding protein YgfZ, folate/pteridine transporter, folate/biopterin transporter, reduced folate carrier family protein, folate/methotrexate transporter FT1, Folate Transporters. A total of 234 proteins identified were involved in the transportation of folate across 63 stains, 23 pathogen species and 12 phyla. Of these, 7% were localized on the mitochondria, with 15% possessing signal peptides. Evolutionary phylogenetic analyses point to the similarities of the identified proteins. Our analysis reveals the presence of folate and folate salvage transporters in a wide diversity of pathogens offering novel possibilities for potential drug development targeting folate transporter routes for global infectious disease control. Keywords: Folate transporter; Eukaryotic pathogens; Drug discovery; Putative homologues.

29. A Systematic Solution to Map Processed Data in tranSMART to Raw Data in Multiple Repositories

Chao Zhang, VU Amsterdam, The Netherlands

Background: With the evolving of high-throughput experimental techniques, large amounts of molecular profiling data are becoming available for regular clinical studies. These data need to be stored, processed, archived, distributed and, more importantly, linked. In ELIXIR pilot project, we focus on connecting the archival storage of such data, with databases that store the processed data and visualised workflow systems that manage the computational pipelines. After perusing the processed data, users often come back to the raw data not only to reconfirm the data processing but also to further explore the raw data in workflow systems like Galaxy. In the Translational Research IT (TraIT) project of Center for Translational Molecular Medicine, raw data are stored in European Genome-phenome Archive (EGA); and processed data, tranSMART. In light of the distinct interpretations of data ontology structures in different repositories, we are aiming to establish the a systematic, flexible and sustainable mapping between the processed data and the corresponding raw data regardless of their repositories. Description: We propose a flexible and systematic scheme by introducing an ontology structure with a few stable mapping concepts -- Study, Sample Derived Experimental Data Unit(SDEDU) and File -- to connect different repositories, separating the three essential processes -- data uploading, data ontology structuring and data retrieving -- in such a way that a general upload and structuring methodology can be used for malleable data retrieving. To demonstrate this ontology structure, we create a RDF graph demo using CTMM-TraIT Cell Line Use Case (CLUC) data. Conclusion: This ontology structure can be extended safely by adding any extra ontology structures to meet the additional requirements of the study. By clustering the files related, the introduced mapping concept SDEDU bridges the gap between study and file level where the interpretation of structures in-between alway differs system-to-system. Moreover, stable location identifiers for these concepts, enables us to retrieve the data both in the short term via pragmatic solutions and in the longer term by using automated access.

30. Hybrid sequencing approach identifies 'Translocatable Units', a circular DNA molecule generated by an intramolecular replicative transposition

Hiren Ghosh, Institute of Medical Microbiology, Justus Liebig University Giessen and German Center for Infection Research (DZIF), Partner site Giessen-Marburg-Langen, Giessen, Germany

A circular DNA molecule generated by an intramolecular replicative transposition comprising of only a single copy of IS26 along with an adjacent DNA segment has recently been identified and termed a'Translocatable Unit' (TU). TUs are genetic elements, capable of generating tandem arrays of antibiotic resistance genes by a mechanism distinct from transposition. Analysis of whole genomes by hybrid sequencing approach allowed us to identify extraneous contigs i.e. non-chromosomal or plasmid, termed as TUs. The whole genome of the E. coli strain V282 was determined using short reads (Illumina) and long reads (PacBio), by mapping and assembly by RS HGAP . Several circular contigs could not be unambiguously assigned to either plasmid or the chromosome. We examined these contigs for their sequence, genes encoded, nucleotide identity and genetic architecture. A single circular contig of 5.11 Mb was identified as the chromosome and an additional circular contig of 118.3 kb, designated as the IncFIA-FII plasmid pECOV282. In addition, two discrete circular contigs of 16.3 kb and 23.7 kb in size, harboring the blaCTX-M-15 allele were detected. High-confidence assembly eliminated the possibility of these two contigs being tandemly duplicated on pECOV282. Re-assessment of these two circular contigs revealed that they represent intramolecular replicative elements that are generated from the plasmid and comprise of IS26 elements with adjacent DNA segments which we designated as TU-1 and TU-2 respectively. TU-1 is a circular DNA element carrying multiple copies of IS26 and antibiotic resistance genes such as ant1, blaCTX-M-15, the major facilitator superfamily (MFS) efflux pump transporters mphA, emrE, srpC and mdfA, genes involved in folate biosynthesis (folA and folP), the phenolic acid stress response regulator padR, and cyclic-di-GMP phosphodiesterase adrB genes . TU-2 additionally carries the antibiotic resistance genes aacA4, catB3 and blaOXA-1. TUs plays a significant role in mobilization of antibiotic resistance genes. Accurate in-depth sequencing of a bacterial isolate provides information not only on the 'static' genome but also captures dynamic changes mediated by mobile genetic elements. These studies can be extended to create a catalogue of elements reflecting dynamic processes involving replication, recombination and reorganization of bacterial genomes and their accessory elements.

31. Evalution of the Hypoglycemic and Antioxidant Potentials of Aqueous and Ethanolic Corm Extracts of Baka (Gladiolus psittacinus) in Diabetic Rats

Oluseyi Akinloye, Federal University of AgrIculture, Abeokuta

Diabetes mellitus is a metabolic disorder affecting all age groups. Management of diabetes without side effects is still a challenge in medicine. Baka (Gladiolus psittacinus) is used by traditional healers in Southwest Nigeria as a recipe for treating diabetes mellitus. This study evaluated the effects of oral administration of two doses (100 and 200 mg/kg) of aqueous and ethanolic extracts of G. psittacinus corm on blood glucose level. Hepatic enzymic and non-enzymic anti-oxidants as well as liver and kidney functions in diabetic rats were evaluated. Fifty-six rats, (150-220g) were randomly divided into seven groups of eight animals each: non-diabetic control group, diabetic control group, aqueous extract-treated groups (100 and 200 mg/kg), ethanolic extract-treated groups (100 and 200 mg/kg) and diabetic-treated group (1mg/kg Glibenclamide- a standard anti-diabetic drug). Diabetes was induced with 50mg/kg streptozotocin. The animals were administered the extracts daily for 21 days orally. Four animals from each group were sacrificed on the eleventh day after treatment and the rest, on the twenty-first day after overnight fast. Blood samples were collected for glucose and protein determinations. Liver samples were taken for histopathological examination. Homogenate (10% w/v) of liver was used for determination of glutathione, vitamin C and malondialdehyde levels and activities of catalase, superoxide dismutase and glutathione-transferase. Data collected were analyzed using ANOVA and means were separated using DMRT. The results showed that aqueous and ethanolic extracts of G. psittacinus corm (100 and 200 mg/kg) significantly ($p<0.05$) reduced blood glucose by 39.6%, 40.4%, 95.3% and 79.53% and lipid peroxidation by 51.2%, 36.9%, 46.2% and 40.6% respectively relative to diabetic control group. Levels of protein, urea, creatinine and bilirubin in plasma were significantly ($p<0.05$) reduced in the extract-treated groups compared to the diabetic control group after twenty-one days of treatment. Antioxidant activities of catalase, superoxide dismutase and glutathione-transferase improved significantly ($p<0.05$) after treatment with extracts. The results indicated that the aqueous and ethanolic extracts of Gladiolus psittacinus corm have the potentials to correct hyperglycemia in diabetes, reverse some metabolic abnormalities (decreased antioxidants levels, increased plasma protein, bilirubin and a creatinine concentrations) associated with diabetic condition and peroxidative damage to the liver.

32. Clinical barriers for the application of Pharmacogenetics in the UK: Statistical analysis of limitations of pharmacogenetic resources with a focus on Warfarin and Clopidogrel treatment and impact of the decreasing costs of Pharmacogenetic tests

Shweta Joshi, University College London, United Kingdom

Background and Hypothesis A large amount of evidence is available to physicians in UK about pharmacogenetic tests for prescription of Warfarin and Clopidogrel. The resources available to support the cardiologists on how to and when to use pharmacogenetic tests need to be more accessible. The lack of clear guidance and high perceived costs of the pharmacogenetics tests act as barriers to adoption of testing before prescription. The principal aims of this study are: 1. To understand the opinion of cardiologists about the available resources that provide guidance on pharmacogenetic testing with respect to Warfarin and Clopidogrel. 2. To determine the limitations of the current resources. 3. To access the awareness about the decreasing cost of the pharmacogenetic tests. Description The design of the project and the methodology to be followed, is based on a study carried out in the United States in 2014. The methodology that will be followed in the project can be explained in the following steps: 1. Questionnaire A questionnaire (15-20 questions) long will be designed so that information about the knowledge gap, limitations of the resources and the decreasing costs of pharmacogenetic testing can be recorded. 2. Distribution As the study will focus on pharmacogenetics pertaining to warfarin and clopidogrel. Cardiologists will be included in the sample being studied. 3. Analysis The data generated will be analysed using non-parametric tests like Mann-Whitney and Kruskal-Wallis. This will be used to identify statistically significant relations in the data. Main Findings and Project Impact This project will make an important assessment about the impact of the support mechanisms available for the interpretation of pharmacogenetic tests for warfarin and clopidogrel and perceived costs associated with genetic testing. The results of the study will be used to make some inferences about the improvements that are needed in the interface between the cardiologists and pharmacogenetic testing. A pilot study has been carried out at the annual conference of the British Cardiovascular Society and the final leg of the study is being carried out right now. The results for the same will be presented at the conference.

33. Knowledge extraction for the life sciences

Umesh Nandal, Elsevier, The Netherlands

Drug discovery and development is a long and complex process with the challenges of high costs and high failure rates. High failure rates can be largely attributed to improper selection of the candidate drug target. In order to formulate hypotheses and design experiments for the identification of the lead drug targets, academic and pharmaceutical research rely heavily on peer-reviewed scientific literature for the extraction of high-quality information. With the advancement in technologies, peer-reviewed articles and patent literature are growing exponentially in numbers. Manual extraction and retrieval of high-quality information from this wealth of unstructured data is extremely time consuming and expensive. Elsevier overcomes these limitations by applying natural language processing (NLP) techniques on unstructured text documents such as document classification, named entity recognition and relation extraction, using machine learning approaches. Elsevier provides high-quality solutions to support and accelerate the processes at the early stage of drug discovery and development. Elsevier provides solutions such as (1) Reaxys; that delivers experimental facts on chemical structures, properties, reactions, compound-target affinity data and pharmacokinetics, (2) PharmaPendium; a database that provide fully searchable FDA/EMA drug approval documents and comparative extracted drug safety data, (3) Embase; a comprehensive biomedical literature database, and, (4) PathwayStudio; that enables better understanding of biological processes underlying disease progression and treatment response.

34. A vicious cycle in mammalian fatty-acid oxidation

Anne-Claire Martines, University Medical Centre Groningen, The Netherlands

Mitochondrial fatty-acid beta-oxidation (mFAO) plays a key role in diseases of energy metabolism. Its complex kinetic structure contains multiple reversible reaction cycles, with reactants of different carbon-chain lengths competing for promiscuous enzymes, and intermediate metabolites esterified to a limited coenzyme A (CoA) pool. In patients with a deficiency in the mFAO enzyme MCAD, fasting combined with fever can cause a sudden life-threatening drop in blood glucose levels. In our experimentally validated dynamic mFAO model we observed flux decline at high substrate concentrations. This was aggravated in the MCAD-knockout model, possibly explaining the glucose levels drop in patients. To elucidate the mechanism underlying the flux decline and to identify novel rescue mechanisms that explain the phenotypic heterogeneity in MCAD deficiency, we applied metabolic control analysis, quantified the flux regulation by each variable metabolite in the pathway, and analyzed the transcriptome of wild-type and MCAD-knockout mouse liver. Upon flux decline, the last enzyme of the pathway, MCKAT, gained flux control. Furthermore, MCKAT's promiscuity, thus its ability to catalyze multiple reactions, was both necessary and sufficient to elicit the flux decline. Quantification of regulation by internal metabolites revealed a vicious cycle: Since the acyl-CoA substrate of each mFAO cycle is also the product of the previous cycle, it provides substrate to MCKAT and inhibits it at the same time. This leads to accumulation of MCKAT substrates, which propagates to preceding CoA ester intermediates in the cycle. Consequently, free CoA levels drop. Since CoA is a co-substrate of MCKAT, this limits its activity further, completing the vicious cycle. As MCAD-deficient patients are healthy under most conditions we hypothesized that surrounding pathways, including acyl-CoA-thioesterase-catalyzed reactions modulate the mFAO flux. Addition of a thioesterase to the model indeed attenuated flux decline, particularly in the MCAD-knockout model. This may be a protective mechanism. Notably, fasting induced a 2-fold higher upregulation of thioesterase expression in MCAD-knockout - compared wild type mouse liver. We are currently testing our predictions in an MCAD-knockout of the human HepG2 cell line, and are extending the model with other relevant pathways, to gain more insight into the loss of robustness in mFAO diseases.